

Современные методы аспектно-ориентированного анализа эмоциональной окраски*

¹И. А. Андрианов <ivan.andrianov@ispras.ru>

¹В. Д. Майоров <vmayorov@ispras.ru>

^{1,2,3}Д. Ю. Турдаков <turdakov@ispras.ru>

¹ Институт системного программирования РАН,
109004, Россия, г. Москва, ул. А. Солженицына, дом 25

² Московский государственный университет имени М.В. Ломоносова,
119991, Россия, Москва, Ленинские горы, д. 1.

³ Национальный исследовательский университет «Высшая школа экономики»
101000, Россия, Москва, ул. Мясницкая, д.20

Аннотация. Данная работа посвящена обзору методов решения актуальной на сегодняшний день задачи аспектно-ориентированного анализа эмоциональной окраски текстов. Данная задача решалась в рамках нескольких конференций, посвященных автоматическому анализу текстов на естественном языке. Организаторы конференций предлагали участникам площадки для сравнительного тестирования методов. В рамках данной работы рассмотрены методы решения задачи аспектно-ориентированного анализа эмоциональной окраски, предложенные участниками двух таких международных площадок: SemEval-2015 и SentiRuEval-2015.

Ключевые слова: анализ эмоциональной окраски; извлечение аспектных терминов; обработка текстов на естественном языке; машинное обучение

DOI: 10.15514/ISPRAS-2015-27(5)-1

Для цитирования: Андрианов И.А., Майоров В.Д., Турдаков Д.Ю. Современные методы аспектно-ориентированного анализа эмоциональной окраски. Труды ИСП РАН, том 27, вып. 5, 2015 г., стр. 5-22. DOI: 10.15514/ISPRAS-2015-27(5)-1.

1. Введение

Последние годы ознаменовались бурным ростом торговых площадок в интернете. Потребители при выборе товара рассматривают не только описание его фактических характеристик, но и отзывы других потребителей.

* Работа поддержана грантом РФФИ 15-37-20375 мол_а_вед

Производители товаров также заинтересованы в получении информации об успешности продукта от потребителей.

Это приводит к появлению все большего числа площадок для размещения отзывов в интернете. Помимо собственно возможности оставить отзыв в свободной форме, подобные площадки нередко предоставляют и возможность оценить различные аспекты товара (например, качество исполнения или надежность) по какой-либо шкале.

В связи с этим приобрела актуальность задача аспектно-ориентированного анализа эмоциональной окраски. Задача предполагает анализ текстов отзывов с целью автоматического извлечения из них информации об эмоциональной окраске тех или иных аспектов товара.

Актуальность задачи привела к росту числа площадок для сравнительного тестирования подходов, организаторы которых предлагают участникам разработать системы, решающие подзадачи аспектно-ориентированного анализа эмоциональной окраски. При этом организаторы вначале предоставляют участникам тренировочные данные, а затем, непосредственно перед проведением тестирования, данные для тестирования. Примерами таких площадок могут служить: SemEval, проводимый в рамках сообщества ACL, а также SentiRuEval, проводимый в рамках конференции “Диалог”.

В рамках данной статьи рассматриваются методы решения задачи аспектно-ориентированного анализа эмоциональной окраски, предложенные участниками площадок SemEval-2015 [1] и SentiRuEval-2015 [2].

В рамках SemEval-2015 предлагалась задача аспектно-ориентированного анализа эмоциональной окраски (задача 12). Эта задача была разделена на две независимые подзадачи, первая из которых была сформулирована как аспектно-ориентированный анализ эмоциональной окраски внутри предметной области (и обучающая, и тестовая коллекции отзывов относились к одной предметной области), вторая - между предметными областями (обучающая коллекция относилась к предметным областям “автомобили” и “рестораны”, тестовая коллекция относилась к неизвестной заранее предметной области - “отели”). Первая подзадача делилась в свою очередь на секции. Одна из секций была посвящена извлечению аспектных терминов, то есть упоминаний целевых объектов и их характеристик, другая - определению полярности эмоциональной окраски для каждого аспектного термина.

В рамках SentiRuEval-2015 задача аспектно-ориентированного анализа эмоциональной окраски рассматривалась только внутри предметной области и разделялась на секции похожим образом: две подзадачи касались извлечения аспектных терминов, еще одна была посвящена определению полярности эмоциональной окраски.

2. SemEval-2015: Извлечение аспектных терминов

2.1 Описание задачи

Задача извлечения аспектных терминов предлагалась участникам SemEval-2015 как задача 12.1.2 [1]. В качестве входных данных в данной задаче выступали тексты отзывов. На выходе требовалось выявить аспектные термины и их аспекты.

Аспектными терминами (Opinion Target Expression) считались упоминания объекта, для которого написан отзыв, или его аспекта. Например, в предложении “*The pizza was delicious.*” имеется аспектный термин “*pizza*” для аспекта *FOOD#QUALITY*. Упоминания, касающиеся других объектов (например, при сравнении двух объектов) не рассматривались.

Участникам были предоставлены тренировочные данные на английском языке для одной предметной области: 254 размеченных отзыва для ресторанов. Тестовые данные, предоставленные позднее, содержали 96 размеченных отзывов для ресторанов. Балансировка данных по аспектам аспектных терминов не производилась.

2.2 Методы, предложенные участниками

Методы решения задачи извлечения аспектных терминов, предложенные участниками SemEval-2015, можно разбить на 3 категории: методы, основанные на разметке последовательности (использовались большинством участников); методы, основанные на выявлении предметно-специфичной терминологии; методы обучения без учителя.

2.2.1 Методы, основанные на разметке последовательности

Разметка последовательности - это подход, хорошо зарекомендовавший себя в задачах обработки текста, например, NERC [3]. Идея подхода во многом схожа с подходом, основанном на модели классификации, и заключается в том, чтобы построить признаковое описание предложения как последовательность признаковых описаний слов, составляющих предложение. Далее производится обучение модели путем подачи ей на вход меток слов предложений и их признаковых описаний. При обработке новых предложений модель ставит их признаковым описаниям в соответствие наиболее вероятную последовательность меток.

В рамках SemEval-2015 участники применяли как одну из наиболее популярных моделей разметки последовательности CRF [4], так и различные эмуляции разметки последовательности с помощью последовательной классификации слов часто применяемыми в обработке текстов моделями линейного SVM [5] и перцептроном [6] с применением дополнительных признаков: меток, присвоенных классификатором предыдущим словам.

Необходимо отметить, что, как и в случае задач NERC и других задач обработки текста, при решении задачи извлечения аспектных терминов с помощью разметки последовательности важную роль играет кодирование. Кодирование обеспечивает возможность различить два подряд идущих термина с одинаковыми аспектами от одного “длинного” термина. Все участники SemEval-2015, применявшие модели разметки последовательности, применяли BIO-кодирование [7], предполагающее наличие трех видов меток: *B-aspect* (начало аспектного термина для аспекта *aspect*), *I-aspect* (середина/конец аспектного термина для аспекта *aspect*), *O* (отсутствие аспектного термина).

Участниками был предложен широкий набор признаков слов, которые можно разделить на следующие категории: лексические и морфологические признаки, признаки на кластерах слов, синтаксические признаки, признаки именованных сущностей.

Лексические и морфологические признаки, использовавшиеся в том или ином виде большинством участников, включали в себя: слова [8], [9], [10]; части речи слов [10], [11], [12]; леммы слов [10], [11]; префиксы / суффиксы слов [8], [11]; наличие в слове прописных, строчных букв, а также цифр [8], [10], [11]; маски слов (замена символов слова на метки, такие как “строчная буква”, “цифра” и т.п.) [11].

Признаки на кластерах слов основывались на результатах кластеризации слов в неразмеченном корпусе текстов. В качестве корпуса участники применяли Википедию [8], а также автоматически собранные из Web отзывы для соответствующей предметной области [8], [9]. Участники применяли следующие подходы для кластеризации: Brown-кластеры [8], [9], [10], [13]; K-means-кластеры векторов word2vec [8], [9], [14], [15]; а также Clark-кластеры [8], [16].

Синтаксические признаки, использовавшиеся участниками, варьировались от простых: стоит ли слово в начале предложения [8], тип фрагмента (например, “именная фраза”) [11], в который входит слово, до более сложных, основанных на синтаксическом разборе предложений по грамматике зависимостей [17]: головное слово [9], метка входящего ребра [10], метка первого исходящего ребра [10].

Признаки именованных сущностей представляли собой BIO-разметку именованных сущностей [11].

2.2.2 Методы, основанные на выявлении предметно-специфичной терминологии

Методы, основанные на выявлении предметно-специфичной терминологии, строятся на предположении о том, что аспектные термины являются предметно-специфичными терминами [18]. Идея методов состоит в выявлении предметно-специфичных терминов каким-либо из известных методов и построении модели классификации терминов по аспектам.

В качестве модели классификации участники применяли SVM. В качестве признаков для данной модели - вхождение в соответствующие целевой предметной области (выбраны вручную) категории DBPedia [19], число гипонимов / гиперонимов по WordNet [20], число терминов, совместно встречающихся с целевым термином, в Brown-кластерах, а также в тренировочных данных, слова предложения, в которое входит термин.

2.2.3 Методы обучения без учителя

Методы обучения без учителя используют для обучения коллекции неразмеченных аспектными терминами отзывов. Важными преимуществами подобных методов являются возможность автоматического сбора большого набора отзывов для целевой предметной области из Веб и отсутствие необходимости привлекать экспертов для трудоемкой разметки отзывов.

Одним из участников был предложен метод, основанный на графах [21]. Вначале осуществлялась выборка всех существительных из тренировочных данных как кандидатов в аспектные термины, а также прилагательных как кандидатов в оценочные слова.

Далее строился граф, в котором узлами выступали выбранные слова двух типов. Ребрам в данном графе назначались веса по следующей схеме: между однотипными узлами - косинусная близость [22] между векторами word2vec, соответствующими словам-узлам; между разнотипными узлами - частота нахождения слов-узлов в тренировочных данных в одном из выбранных вручную синтаксических отношений.

Узлы-кандидаты в аспектные термины, имевшие наибольшие значения меры PageRank [23], вычисленной на построенном графе, считались аспектными терминами и размечались в тестовых данных путем поиска совпадений по леммам. Аспектом всех аспектных терминов предложения считался аспект, для которого сумма косинусных близостей векторов word2vec вручную выбранных слов и слов предложения была наибольшей.

У данного метода есть существенный недостаток: он не может находить аспектные термины, содержащие более одного слова. Для устранения данного недостатка было предложено выявить самые устойчивые словосочетания в тренировочных данных с помощью меры Log-Likelihood Ratio [24] и рассматривать их как одно слово при обработке.

2.3 Анализ результатов

Для оценки результатов организаторы использовали F1-меру. В качестве базового метода использовался метод, который собирал словарь аспектных терминов по обучающей коллекции, и искал в тестовой коллекции аспектные термины по полному совпадению со словарными.

Базовый метод получил оценку 48,06%, методы участников - оценки от 33,86% до 70,05%. Лучшие результаты показали методы на основе разметки

последовательности, большинство из них превзошло результаты базового метода. Лучшие результаты среди методов на основе разметки последовательности показали методы, использовавшие признаки на кластерах слов. Важными для показа высоких результатов оказались также лексические и морфологические признаки. Методы на основе выявления предметно-специфичной терминологии и методы обучения без учителя показали результаты, схожие с результатами базового метода: от 45,67% до 49,97%.

3. SentiRuEval-2015: Извлечение аспектных терминов

3.1 Описание задачи

Задача извлечения аспектных терминов предлагалась участникам SentiRuEval-2015 как задача А (явные) и задача В (явные, неявные и факты) [2]. В качестве входных данных в данной задаче выступали тексты отзывов. На выходе требовалось выявить аспектные термины и их аспекты.

Явными аспектными терминами назывались термины, содержащие упоминание объекта, для которого написан отзыв, или его аспекта. Неявными аспектными терминами назывались термины, включающие оценочные слова, специфичные для какого-то аспекта: например, “комфортный” представляет позитивную оценку аспекта *интерьер*. Эмоциональными фактами назывались термины, которые представляют объективную информацию, но неявно выражают мнение автора отзыва: например, “была вежлива” представляет позитивную оценку аспекта *обслуживание*.

Участникам были предоставлены тренировочные коллекции на русском языке для двух предметных областей: 201 размеченный отзыв для ресторанов, 217 - для автомобилей. Тестовые данные, предоставленные позднее, содержали 203 отзыва и 201 соответственно. Балансировка данных по аспектам и типам (явные, неявные, факты) аспектных терминов не производилась. Число явных аспектных терминов превышало число неявных/фактов в 4 - 6 раз.

3.2 Методы, предложенные участниками

Участники SentiRuEval-2015 так же, как и участники SemEval-2015, в основном использовали подходы, основанные на разметке последовательности. Так, Рубцова Ю. В. и Кошельников С. А. [25] предложили применять для решения задачи извлечения аспектных терминов CRF с использованием лексических и морфологических признаков (словоформы предыдущего и следующего слова, часть речи и лемма слова). Для задачи А было предложено обучать классификатор на трех классах (В, I, О), а для задачи В, в которой необходимо определять аспектные термины трех типов, было предложено использовать использовать 7 классов (В и I для каждого из трех типов аспектных терминов и класс О).

Майоров В. Д. и др. [26] предложили использовать линейный SVM с дополнительными признаками - метками классов предыдущих слов. Также ими был предложен намного более обширный список использованных признаков классификации, в том числе лексические, морфологические, синтаксические признаки, признаки, кластеры слов (кластеры векторных представлений слов, тематическое моделирование) и признаки, использующиеся в задаче извлечения терминологии [27]. В отличие от решения, предложенного Рубцовой, авторы метода предлагают использовать оригинальный метод (решение задачи А) независимо для каждого из типов аспектов для решения задачи В.

Блинов П. Д. и Котельников Е. В. [28] предложили альтернативный метод решения задачи. Идея метода заключается в использовании семантической близости между словами для определения аспектных терминов. Каждому слову из тестовой выборки ставился в соответствие вектор word2vec и вычислялась косинусная мера близости к словам, которые вошли в состав аспектных терминов в обучающей выборке. Если результат оказывался выше экспериментально определенного порога, то слово помечалось аспектным термином. Для определения многословных аспектных терминов авторами был предложен набор правил для объединения рядом стоящих аспектных терминов, например, найденные однословные аспектные термины объединялись в один многословный, если были разделены предлогом.

Также был предложен метод, основанный на использовании рекуррентных нейронных сетей [29]. Авторами метода были предложены различные варианты решения задачи А: рекуррентная нейронная сеть Эльмана [30], двунаправленная рекуррентная нейронная сеть [31], использовалась длинная короткая память [32]. Для преобразования входных слов в сигналы применяется языковая модель рекуррентных нейронных сетей [33]. Обучение нейронных сетей всех типов проводится с использованием метода обратного распространения ошибки [34]. При обучении, в исходных данных все числа были заменены на специальный маркер. Дополнительно было предложено добавлять во входные данные небольшой Гауссовский шум. Лучшие результаты, согласно данным авторов статьи, показал метод, основанный на двунаправленных рекуррентных нейронных сетях с использованием длинной короткой памяти.

3.3 Анализ результатов

При оценке результатов организаторы использовали F1-меру с макро-усреднением 2 видов: для полного и для частичного совпадения аспектных терминов. В качестве базового метода использовался метод, который размечал как аспектный термин для заданного аспекта слова и словосочетания, которые выступали в качестве аспектных терминов для данного аспекта в тренировочных данных.

Предложенные участниками решения задачи А не смогли показать результаты, значительно превышающие показатели базового метода. При вычислении качества по точному совпадению аспектных терминов лучшим решением оказался метод, основанный на разметке последовательности слов с использованием SVM. Он получил для предметных областей “рестораны” и “автомобили” оценки 63.1% и 67.6% соответственно, в то время как базовый метод получил оценки 60.8% и 59.4%. При учете частичного совпадения аспектных терминов для предметной области “рестораны” лучшими методами оказались подход на основе семантической близости слов (72.8%) и метод на основе рекуррентных нейронных сетей (71.9%). Базовый метод получил оценку 66.5%. Для предметной области “автомобили” лучшие результаты показали метод на основе рекуррентных нейронных сетей (74.8%) и метод, использующий SVM с большим числом лексических, морфологических, синтаксических и семантических признаков (73.0%). Базовый метод получил оценку 69.6%.

Лучшие решения задачи В также незначительно превосходят показатели базового метода. Для обеих предметных областей (“рестораны” и “автомобили”) по точному совпадению аспектных терминов только два метода смогли преодолеть базовый (58.7% и 58.9%): метод на основе нейронных сетей (60.0%, 63.0%) и метод на основе SVM (59.6% и 63.6%). По частичному совпадению аспектных терминов базовый метод (61.9% и 67.4%) смог превзойти лишь метод на основе рекуррентных нейронных сетей (66.8% и 71.5%).

4. SemEval-2015: Анализ полярности аспектных терминов

4.1 Описание задачи

Задача аспектно-ориентированного анализа полярности предложений предлагалась участникам SemEval-2015 как задача 12.1.3 (для известных заранее предметных областей) и задача 12.2 (для неизвестной заранее предметной области) [1]. В качестве входных данных в данной задаче выступали тексты отзывов, у которых каждое предложение было размечено множеством аспектов и аспектных терминов, в отношении которых в предложении выражаются мнения. На выходе требовалось выявить полярность для каждого из указанных аспектных терминов.

Рассматривалось 3 класса полярности: позитивная, негативная и нейтральная. Одновременное наличие как позитивной, так и негативной полярности рассматривалось как нейтральная.

Участникам были предоставлены тренировочные данные на английском языке для двух предметных областей: 254 размеченных отзыва для ресторанов, 277 - для ноутбуков. Тестовые данные, предоставленные позднее, содержали 96 размеченных отзывов для ресторанов, 173 - для ноутбуков, а также 30 - для неизвестной заранее предметной области (отелей). Балансировка данных по

классам полярности не производилась. Позитивный класс был мажорным с относительной частотой вхождений от 53% до 73% в зависимости от предметной области. Нейтральный класс был самым малочисленным с относительной частотой вхождений от 3% до 9%.

4.2 Методы, предложенные участниками

Методы решения задачи аспектно-ориентированного анализа полярности предложений, предложенные участниками SemEval-2015, можно разбить на 2 категории: методы, основанные на классификации; методы обучения без учителя.

4.2.1 Методы, основанные на классификации

Идея методов, основанных на классификации, состоит в выборе контекста аспектного термина и построении модели классификации выбранного контекста по полярности.

В простейшем случае участники выбирали все предложение, в которое входит аспектный термин, в качестве контекста [10], [18]. Такой подход обладает серьезным недостатком: все аспектные термины в предложении получают одинаковую полярность, что в общем случае неверно. Например, в предложении *“I like the somosas, chai, and the chole, but the dhosas and dhal were kinda dissapointing.”* термины *“somosas”* и *“dhosas”* имеют положительную и отрицательную полярности соответственно. Для устранения данного недостатка некоторые участники применяли разбиение предложений по пунктуации [11]. Также применялся и типовой для задач обработки текста способ выбора контекста: окно слов слева-справа [8] и его модификация, использующая не позиции в тексте, а позиции в синтаксическом дереве [35], [36]. Кроме того, применялся подход, выбирающий из предложения и соседних с ним предложений слова с наивысшими весами, например, tf-idf [36], [37].

В качестве моделей классификации участники применяли линейный SVM, логистическую регрессию [38]. Признаки для моделей классификации, предложенные участниками можно разделить на следующие категории: лексические, морфологические и синтаксические признаки, признаки на основе внешних источников данных.

Лексические, морфологические и синтаксические признаки включают в себя: униграммы [10], [11], [18], [35], [36]; биграммы [11], [36]; части речи [8], [35], [36]; леммы [8]; наличие повторяющихся букв [10] и восклицательных знаков [10], [36]; вхождение слова в список вопросительных или условных слов [10]; метки входящих ребер синтаксического дерева [35], [36].

Наиболее часто в качестве внешнего источника данных участники задействовали как составленные вручную, так и собранные автоматически, лексиконы оценочных слов: Bing Liu [8], [11], [18], [35], [36]; General Inquirer

[8], [18], [36]; SentiWordNet [8], [11], [35], [36]; MPQA [11], [35], [36]; Sentiment140 [11], [36]; NRC Hashtag [11], [36]. В качестве признаков на основе лексиконов выступали: полярности, суммы полярностей, максимумы / минимумы полярностей, относительное число позитивных / негативных полярностей. Некоторые участники инвертировали полярность слова, если рядом с ним встречалось слово-отрицание [11], [18], [35], [36]. Помимо лексиконов оценочных слов участники применяли признаки на основе WordNet [10] и Brown-кластеров [11].

4.2.2 Методы обучения без учителя

Одним из участников был предложен следующий метод обучения без учителя [21]. Были вручную выбраны слова, имеющие заданную полярность в любой предметной области: например, “*excellent*”. Далее из тренировочных данных были извлечены слова, имеющие самые высокие значения косинусной близости векторов word2vec с выбранными вручную словами. Нахождение подобных слов рядом с аспектным термином считалось признаком соответствующей полярности аспектного термина. Кроме того, во внимание принимались и слова-отрицания, их наличие рядом с аспектным термином инвертировало его полярность.

4.3 Анализ результатов

При оценке результатов организаторы использовали достоверность, т.е. число аспектов с верно определенной полярностью к общему числу аспектов в предложениях тестовых данных. В качестве базового метода использовался линейный SVM с признаками “мешок слов” и “номер аспекта” (всем аспектам были предоставлены уникальные номера).

Для предметной области ресторанов базовый метод получил оценку 63,55%, методы участников - оценки от 60,71% до 78,69%. Для предметной области ноутбуков базовый метод получил оценку 69,96%, методы участников - оценки от 51,84% до 79,34%.

Лучшими методами на основе классификации стали методы, которые использовали различные оценки слов контекста аспектного термина на основе лексиконов оценочных слов, а также учитывали слова-отрицания. Лучшими подходами к выбору слов контекста аспектного термина стали выбор слов с наивысшим tf-idf в текущем и соседних предложениях, а также разбиение предложения по пунктуации. Методы обучения без учителя показали результаты, близкие к результатам базового метода: от 68,38% до 69,46% в зависимости от предметной области.

Необходимо отметить, что практически все методы, включая базовый, показали для предметной области “ноутбуки” более высокий результат, нежели для предметной области “рестораны”. Это можно объяснить тем, что тестовые данные для предметной области “рестораны” содержали 53,72% аспектных терминов с позитивной полярностью, в то время как

тренировочные – 72,43%. Для предметной области “ноутбуки” же данные показатели – 57% и 55,87% соответственно. Однако чем выше был результат метода, тем меньшей становилась разница результатов для двух предметных областей. Для лучших методов разница результатов была менее 1%, для базового – более 6%.

5. SentiRuEval-2015: Анализ полярности аспектных терминов

5.1 Описание задачи

Задача анализа полярности аспектных терминов предлагалась участникам SentiRuEval-2015 как задача C [2]. В качестве входных данных в данной задаче выступали тексты отзывов, размеченные аспектными терминами и их аспектами. На выходе требовалось выявить полярность аспектных терминов.

Рассматривалось 4 класса полярности: позитивная, негативная, нейтральная, обе. Нейтральная использовалась также и в случае отсутствия эмоциональной окраски: например, в предложении *“Так-же при покупке авто стоит сразу-же надеть чехлы на сиденья так как обивка сидений тоже бюджетная.”*, где мнение выражается в отношении аспектного термина *“обивка сидений”*, но не термина *“чехлы на сиденья”*.

Участникам были предоставлены тренировочные коллекции на русском языке для двух предметных областей: 201 размеченный отзыв для ресторанов, 217 - для автомобилей. Тестовые данные, предоставленные позднее, содержали 203 отзыва и 201 соответственно. Балансировка данных по классам полярности не производилась. Позитивная полярность была мажорным классом для всех предметных областей, вслед за ней располагались негативная и нейтральная, класс “обе” был крайне редким (не более 115 терминов в каждом наборе данных). Отношение числа терминов в позитивном классе к их числу в негативном/нейтральном располагалось в интервале от 3 до 6.

5.2 Методы, предложенные участниками

Участниками были предложены методы на основе word2vec [28] и рекуррентных нейронных сетей [29].

Рассмотрим подробнее метод на основе word2vec, предложенный одним из участников. Вначале были выбраны эталонные слова положительной / отрицательной полярностей для каждой из предметных областей. Затем для каждого прилагательного и глагола из обучающих данных была подсчитана сумма косинусных близостей векторов word2vec данного слова и слов-эталонов. В качестве полярности слова выбиралась та, для которой такая сумма была наибольшей. В случае если перед словом стояло слово отрицание, полярность инвертировалась.

Далее были использованы дополнительные коллекции отзывов соответствующих предметных областей, собранные участником, с оценками по десятибалльной шкале. Коллекции были разбиты по оценкам на резко положительные и резко отрицательные отзывы, на каждом фрагменте была вычислена мера PMI [39] для каждого прилагательного / глагола из обучающих данных.

Вычисленные с помощью word2vec полярности и значения PMI всех прилагательных и глаголов, находящихся рядом с аспектным термином, использовались как признаки для модели классификации Gradient Boosting [40].

Полярность “оба” выбиралась для аспектных терминов, рядом с которыми в рамках предложения находилось слово “но”.

5.3 Анализ результатов

При оценке результатов организаторы использовали F1-меры с макро- и микро-усреднением. В качестве базового метода использовался метод, который назначал аспектным терминам наиболее распространенный класс полярности для их аспекта в тренировочных данных (это был “позитивно” во всех случаях).

По F1-мере с макро-усреднением базовый метод получил оценку 26,7%, лучший из 7 методов участников (Gradient Boosting) - оценку 55,4%. Остальные методы участников либо не смогли превысить оценку базового метода, либо превысили ее незначительно. Схожая картина наблюдалась и в оценках по F1-мере с микро-усреднением: базовый метод получил оценку 71,0%, лучший из методов участников (также Gradient Boosting) - 82,4%.

6. Заключение

Участники международных площадок для сравнительного тестирования методов аспектно-ориентированного анализа эмоциональной окраски в основном предлагали методы, основанные на машинном обучении с учителем. Однако результаты тестирования показали незначительное улучшение качества по сравнению с простыми базовыми методами решения этой задачи. Таким образом, в результате обзора методов, предложенных в рамках SemEval-2015 и SentiRuEval-2015, можно сделать вывод, что задачи автоматического извлечения аспектных терминов и определения полярности эмоциональной окраски все еще далеки от окончательного решения.

Организаторы SemEval-2015 предложили следующий этап сравнительного тестирования в рамках площадки SemEval-2016*, где рассматривается та же задача, но расширен перечень предметных областей, и существенно расширен список целевых языков.

* <http://alt.qcri.org/semeval2016/task5/>

Список литературы

- [1]. Pontiki M., Galanis D., Papageorgiou H., Manandhar S., Androutsopoulos I. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 486–495.
- [2]. Loukachevitch N., Blinov P., Kotelnikov E., Rubtsova Y., Ivanov V., Tutubalina E. SentiRuEval: Testing Object-oriented Sentiment Analysis Systems in Russian. Proceedings of the 21st International Conference on Computational Linguistics Dialog-2015, 2015, volume 2, pp. 12–24.
- [3]. Ratinov L., Roth D. Design challenges and misconceptions in named entity recognition. Proceedings of the Thirteenth Conference on Computational Natural Language Learning, 2009, pp. 147–155.
- [4]. Lafferty J., McCallum A., Pereira F. Conditional random fields: probabilistic models for segmenting and labeling sequence data. Proceedings of the Eighteenth International Conference on Machine Learning, 2001, pp. 282–289.
- [5]. Vapnik V. Statistical Learning Theory. Wiley, New York, NY, 1998.
- [6]. Collins M. Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms. Proceedings of the ACL-02 conference on Empirical methods in natural language processing, 2002, volume 10, pp. 1–8.
- [7]. Ramshaw L. A., Marcus M. P. Text chunking using transformation-based learning. Natural language processing using very large corpora, Springer Netherlands, 1999, pp. 157–176.
- [8]. San Vicente I., Saralegi X., Agerri R. EliXa: A modular and flexible ABSA platform. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 748–752.
- [9]. Toh Z., Su J. NLANGP: Supervised Machine Learning System for Aspect Category Classification and Opinion Target Extraction. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 496–501.
- [10]. Guha S., Joshi A., Varma V. SIEL: Aspect Based Sentiment Analysis in Reviews. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 759–766.
- [11]. Hamdan H., Bellot P., Bechet F. Lsislif: CRF and Logistic Regression for Opinion Target Extraction and Sentiment Polarity Analysis. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 753–758.
- [12]. Koppula A., Pallela R., Repaka R., Movva V. UMDuluth-CS8761-12: A Novel Machine Learning Approach for Aspect Based Sentiment Analysis. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 742–747.
- [13]. Brown P. F., Desouza P. V., Mercer R. L., Pietra V. J. D., Lai J. C. Class-based n-gram models of natural language. Computational linguistics, 1992, volume 18, issue 4, pp. 467–479.
- [14]. Hartigan J. A., Wong M. A. Algorithm AS 136: A k-means clustering algorithm. Applied statistics, 1979, pp. 100–108.
- [15]. Mikolov T., Yih W., Zweig G. Linguistic Regularities in Continuous Space Word Representations. Proceedings of HLT-NAACL, 2013, pp. 746–751.
- [16]. Clark A. Combining distributional and morphological information for part of speech induction. Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics, 2003, volume 1, pp. 59–66.

- [17]. Eisner J. M. Three new probabilistic models for dependency parsing: An exploration. Proceedings of the 16th conference on Computational linguistics, 1996, volume 1, pp. 340-345.
- [18]. De Clercq O., Van de Kauter M., Lefever E. and Hoste V. LT3: Applying Hybrid Terminology Extraction to Aspect-Based Sentiment Analysis. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 719-724.
- [19]. Auer S., Bizer C., Kobilarov G., Lehmann J., Cyganiak R., Ives Z. Dbpedia: A nucleus for a web of open data. Springer Berlin Heidelberg, 2007, pp. 722-735.
- [20]. Miller G. A. WordNet: a lexical database for English. Communications of the ACM, 1995, volume 38, issue 11, pp. 39-41.
- [21]. Garcia-Pablos A., Cuadros M., Rigau G. V3: Unsupervised Aspect Based Sentiment Analysis for SemEval-2015 Task 12. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 714-718.
- [22]. Yates R. B., Neto B. R. Modern Information Retrieval. ADDISON-WESLEY, New York, 1999, pp. 501.
- [23]. Page L., Brin S., Motwani R., Winograd T. The PageRank Citation Ranking: Bringing Order to the Web. Stanford InfoLab, 1999, pp. 17.
- [24]. Owen A. B. Empirical likelihood ratio confidence intervals for a single functional. Biometrika, 1988, volume 75, issue 2, pp. 237-249.
- [25]. Rubtsova Y. V., Koshelnikov S. A. Aspect Extraction Using Conditional Random Fields. <http://www.dialog-21.ru/digests/dialog2015/materials/pdf/RubtsovaYVKoshelnikovSA.pdf>, 2015.
- [26]. Mayorov V., Andrianov I., Astrakhantsev N., Avanesov V., Kozlov I., Turdakov D. A High Precision Method for Aspect Extraction in Russian. Proceedings of the 21st International Conference on Computational Linguistics Dialog-2015, 2015, volume 2, pp. 58-67.
- [27]. Астраханцев Н. А. Автоматическое извлечение терминов из коллекции текстов предметной области с помощью Википедии. Труды ИСП РАН, 2014, том 26, выпуск 4, с. 7-20.
- [28]. Blinov P. D., Kotelnikov E. V. Semantic Similarity for Aspect-Based Sentiment Analysis. Proceedings of the 21st International Conference on Computational Linguistics Dialog-2015, 2015, volume 2, pp. 36-45.
- [29]. Tarasov D. S. Deep Recurrent Neural Networks for Multiple Language Aspect-based Sentiment Analysis of User Reviews. Proceedings of the 21st International Conference on Computational Linguistics Dialog-2015, 2015, volume 2, pp. 77-88.
- [30]. Elman J. Finding structure in time. Cognitive science, 1990, volume 14(2), pp. 179-211.
- [31]. Schuster M., Kuldip K. P. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 1997, volume 45(11), pp. 2673-2681.
- [32]. Hochreiter S., Schmidhuber J.. Long short-term memory. Neural computation, 1997, volume 9(8), pp. 1735-1780.
- [33]. Mikolov T., Karafiat M., Burget L., Cernocky J., Khudanpur S. Recurrent neural network based language model. In INTERSPEECH, 2010, pp. 1045-1048.
- [34]. Werbos P. J. Backpropagation through time: what it does and how to do it. Proceedings of the IEEE, 1990, volume 78(10), 1550-1560.
- [35]. Jimenez-Zafra S., Martinez-Camara E., Martin-Valdivia M., Urena-Lopez L. SINAI: Syntactic approach for Aspect Based Sentiment Analysis. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 730-735.

- [36]. Zhang Z., Lan M. ECNU: Extracting Effective Features from Multiple Sequential Sentences for Target-dependent Sentiment Analysis in Reviews. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 736–741.
- [37]. Ramos J. Using tf-idf to determine word relevance in document queries. Proceedings of the first instructional conference on machine learning, 2003, pp. 45–65.
- [38]. Press S. J., Wilson S. Choosing between logistic regression and discriminant analysis. Journal of the American Statistical Association, 1978, volume 73, issue 364, pp. 699–705.
- [39]. Church K. W., Hanks P. Word association norms, mutual information, and lexicography. Computational linguistics, 1990, volume 16, issue 1, pp. 22–29.
- [40]. Friedman J. H. Stochastic gradient boosting. Computational Statistics & Data Analysis, 2002, volume 38, issue 4, pp. 367–378.

Modern Approaches to Aspect-Based Sentiment Analysis

¹I. Andrianov <ivan.andrianov@ispras.ru>

¹V. Mayorov <vmayorov@ispras.ru>

^{1,2,3}D. Turdakov <turdakov@ispras.ru>

¹Institute for System Programming of the RAS,
25, Alexander Solzhenitsyn Str., Moscow, 109004, Russia.

²Lomonosov Moscow State University,
GSP-1, Leninskie Gory, Moscow, 119991, Russia.

³National Research University Higher School of Economics (HSE)
11 Myasnitskaya Ulitsa, Moscow, 101000, Russia

Abstract. The paper presents a survey of methods solving the actual task of aspect-based sentiment analysis. Solutions for this task were proposed at multiple natural language processing conferences. Organizers of these conferences proposed evaluation platforms for methods for aspect-based sentiment analysis. This paper describes methods proposed by participants of two international evaluation platforms: SemEval-2015 focusing on English texts and SentiRuEval-2015 focusing on Russian texts.

SemEval-2015 organizers provided participants with the task 12.2 for English language and two domains: restaurants and laptops. The task was split to multiple subtasks two of which are described in this paper: opinion target expression (both explicit and implicit ones) extraction and sentiment polarity detection. Described methods for opinion target expression can be split to the following categories: sequence labeling, domain-specific terminology extraction and unsupervised learning. Methods for sentiment polarity detection varied from classification-based to unsupervised learning.

SentiRuEval-2015 organizers provided participants with the tasks A, B and C for Russian language and two domains: restaurants and automobiles. Task A was devoted to explicit aspect term extraction, task B – to explicit, implicit and factual aspect term extraction. Sentiment polarity detection was subject of the Task C. Described methods for aspect term

extraction can be classified as following: sequence labeling, word2vec-based and neural network-based. Methods for sentiment polarity detection varied from word2vec-based to neural network-based.

Keywords: sentiment analysis; aspect extraction; text processing; machine learning

DOI: 10.15514/ISPRAS-2015-27(5)-1

For citation: Andrianov I., Mayorov V., Turdakov D. Modern Approaches to Aspect-Based Sentiment Analysis. Trudy ISP RAN/Proc. ISP RAS, vol. 27, issue 5, 2015, pp. 5-22 (in Russian). DOI: 10.15514/ISPRAS-2015-27(5)-1.

References

- [1]. Pontiki M., Galanis D., Papageorgiou H., Manandhar S., Androutsopoulos I. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 486–495.
- [2]. Loukachevitch N., Blinov P., Kotelnikov E., Rubtsova Y., Ivanov V., Tutubalina E. SentiRuEval: Testing Object-oriented Sentiment Analysis Systems in Russian. Proceedings of the 21st International Conference on Computational Linguistics Dialog-2015, 2015, volume 2, pp. 12–24.
- [3]. Ratinov L., Roth D. Design challenges and misconceptions in named entity recognition. Proceedings of the Thirteenth Conference on Computational Natural Language Learning, 2009, pp. 147–155.
- [4]. Lafferty J., McCallum A., Pereira F. Conditional random fields: probabilistic models for segmenting and labeling sequence data. Proceedings of the Eighteenth International Conference on Machine Learning, 2001, pp. 282–289.
- [5]. Vapnik V. Statistical Learning Theory. Wiley, New York, NY, 1998.
- [6]. Collins M. Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms. Proceedings of the ACL-02 conference on Empirical methods in natural language processing, 2002, volume 10, pp. 1–8.
- [7]. Ramshaw L. A., Marcus M. P. Text chunking using transformation-based learning. Natural language processing using very large corpora, Springer Netherlands, 1999, pp. 157-176.
- [8]. San Vicente I., Saralegi X., Agerri R. EliXa: A modular and flexible ABSA platform. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 748–752.
- [9]. Toh Z., Su J. NLANGP: Supervised Machine Learning System for Aspect Category Classification and Opinion Target Extraction. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 496–501.
- [10]. Guha S., Joshi A., Varma V. SIEL: Aspect Based Sentiment Analysis in Reviews. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 759–766.
- [11]. Hamdan H., Bellot P., Bechet F. Lsislif: CRF and Logistic Regression for Opinion Target Extraction and Sentiment Polarity Analysis. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 753–758.
- [12]. Koppula A., Pallela R., Repaka R., Movva V. UMDuluth-CS8761-12: A Novel Machine Learning Approach for Aspect Based Sentiment Analysis. Proceedings of the

- 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 742–747.
- [13]. Brown P. F., Desouza P. V., Mercer R. L., Pietra V. J. D., Lai J. C. Class-based n-gram models of natural language. *Computational linguistics*, 1992, volume 18, issue 4, pp. 467–479.
- [14]. Hartigan J. A., Wong M. A. Algorithm AS 136: A k-means clustering algorithm. *Applied statistics*, 1979, pp. 100-108.
- [15]. Mikolov T., Yih W., Zweig G. Linguistic Regularities in Continuous Space Word Representations. *Proceedings of HLT-NAACL*, 2013, pp. 746-751.
- [16]. Clark A. Combining distributional and morphological information for part of speech induction. *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics*, 2003, volume 1, pp. 59–66.
- [17]. Eisner J. M. Three new probabilistic models for dependency parsing: An exploration. *Proceedings of the 16th conference on Computational linguistics*, 1996, volume 1, pp. 340-345.
- [18]. De Clercq O., Van de Kauter M., Lefever E. and Hoste V. LT3: Applying Hybrid Terminology Extraction to Aspect-Based Sentiment Analysis. *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015)*, 2015, pp. 719–724.
- [19]. Auer S., Bizer C., Kobilarov G., Lehmann J., Cyganiak R., Ives Z. Dbpedia: A nucleus for a web of open data. *Springer Berlin Heidelberg*, 2007, pp. 722-735.
- [20]. Miller G. A. WordNet: a lexical database for English. *Communications of the ACM*, 1995, volume 38, issue 11, pp. 39-41.
- [21]. Garcia-Pablos A., Cuadros M., Rigau G. V3: Unsupervised Aspect Based Sentiment Analysis for SemEval-2015 Task 12. *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015)*, 2015, pp. 714–718.
- [22]. Yates R. B., Neto B. R. *Modern Information Retrieval*. ADDISON-WESLEY, New York, 1999, pp. 501.
- [23]. Page L., Brin S., Motwani R., Winograd T. The PageRank Citation Ranking: Bringing Order to the Web. *Stanford InfoLab*, 1999, pp. 17.
- [24]. Owen A. B. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 1988, volume 75, issue 2, pp. 237-249.
- [25]. Rubtsova Y. V., Koshelnikov S. A. Aspect Extraction Using Conditional Random Fields. <http://www.dialog-21.ru/digests/dialog2015/materials/pdf/RubtsovaYVKoshelnikovSA.pdf>, 2015.
- [26]. Mayorov V., Andrianov I., Astrakhtantsev N., Avanesov V., Kozlov I., Turdakov D. A High Precision Method for Aspect Extraction in Russian. *Proceedings of the 21st International Conference on Computational Linguistics Dialog-2015*, 2015, volume 2, pp. 58–67.
- [27]. Astrakhtantsev N. Avtomaticheskoe izvlechenie terminov iz kolleksii tekstov predmetnoi oblasti s pomoshch'yu Vikipedii [Automatic term acquisition from domain-specific text collection by using Wikipedia]. *Trudy ISP RAN [The Proceedings of ISP RAS]*, 2014, volume 26, issue 4, pp. 7–20.
- [28]. Blinov P. D., Kotelnikov E. V. Semantic Similarity for Aspect-Based Sentiment Analysis. *Proceedings of the 21st International Conference on Computational Linguistics Dialog-2015*, 2015, volume 2, pp. 36–45.
- [29]. Tarasov D. S. Deep Recurrent Neural Networks for Multiple Language Aspect-based Sentiment Analysis of User Reviews. *Proceedings of the 21st International Conference on Computational Linguistics Dialog-2015*, 2015, volume 2, pp. 77–88.
- [30]. Elman J. Finding structure in time. *Cognitive science*, 1990, volume 14(2), pp. 179–211.

- [31]. Schuster M., Kuldip K. P. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 1997, volume 45(11), pp. 2673–2681.
- [32]. Hochreiter S., Schmidhuber J.. Long short-term memory. *Neural computation*, 1997, volume 9(8), pp. 1735 –1780.
- [33]. Mikolov T., Karafiat M., Burget L., Cernocky J., Khudanpur S. Recurrent neural network based language model. In *INTERSPEECH*, 2010, pp. 1045–1048.
- [34]. Werbos P. J. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 1990, volume 78(10), 1550–1560.
- [35]. Jimenez-Zafra S., Martinez-Camara E., Martin-Valdivia M., Urena-Lopez L. SINAI: Syntactic approach for Aspect Based Sentiment Analysis. *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015)*, 2015, pp. 730–735.
- [36]. Zhang Z., Lan M. ECNU: Extracting Effective Features from Multiple Sequential Sentences for Target-dependent Sentiment Analysis in Reviews. *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015)*, 2015, pp. 736–741.
- [37]. Ramos J. Using tf-idf to determine word relevance in document queries. *Proceedings of the first instructional conference on machine learning*, 2003, pp. 45–65.
- [38]. Press S. J., Wilson S. Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association*, 1978, volume 73, issue 364, pp. 699-705.
- [39]. Church K. W., Hanks P. Word association norms, mutual information, and lexicography. *Computational linguistics*, 1990, volume 16, issue 1, pp. 22-29.
- [40]. Friedman J. H. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 2002, volume 38, issue 4, pp. 367-378.